Children's Yale-Brown Obsessive Compulsive Scale Modified for Pervasive Developmental Disorders

LAWRENCE SCAHILL, M.S.N., PH.D., CHRISTOPHER J. MCDOUGLE, M.D., SUSAN K. WILLIAMS, PH.D., ANASTASIA DIMITROPOULOS, PH.D., MICHAEL G. AMAN, PH.D., JAMES T. MCCRACKEN, M.D., ELAINE TIERNEY, M.D., L. EUGENE ARNOLD, M.E.D., M.D., PEGEEN CRONIN, PH.D., MARCO GRADOS, M.D., JASWINDER GHUMAN, M.D., KATHLEEN KOENIG, M.S.N., KRISTEN S.L. LAM, PH.D., JAMES MCGOUGH, M.D., DAVID J. POSEY, M.D., LOUISE RITZ, M.B.A., NAOMI B. SWIEZY, PH.D., AND BENEDETTO VITIELLO, M.D.

ABSTRACT

Objective: To examine the psychometric properties of the Children's Yale-Brown Obsessive Compulsive Scales (CYBOCS) modified for pervasive developmental disorders (PDDs). **Method:** Raters from five Research Units on Pediatric Psychopharmacology (RUPP) Autism Network were trained to reliability. The modified scale (CYBOCS-PDD), which contains only the five Compulsion severity items (range 0–20), was administered to 172 medication-free children (mean 8.2 ± 2.6 years) with PDD (autistic disorder, $n = 152$; Asperger's disorder, $n = 6$; PDD not otherwise specified, $n = 14$) participating in RUPP clinical trials. Reliability was assessed by intraclass correlation coefficient (ICC) and internal consistency by Cronbach's α coefficient. Correlations with ratings of repetitive behavior and disruptive behavior were examined for validity. **Results:** Eleven raters showed excellent reliability (ICC = 0.97). The mean CYBOCS score was 14.4 (± 3.86) with excellent internal consistency ($\alpha = .85$). Correlations with other measures of repetitive behavior ranged from $r = 0.11$ to $r = 0.28$ and were similar to correlations with measures of irritability ($r = 0.24$) and hyperactivity ($r = 0.25$). Children with higher scores on the CYBOCS-PDD had higher levels of maladaptive behaviors and lower adaptive functioning. **Conclusions:** The five-item CYBOCS-PDD is reliable, distinct from other measures of repetitive behavior, and sensitive to change. *J. Am. Acad. Child Adolesc. Psychiatry*. 2006;45(9):1114–1123. **Key Words:** autism, pervasive developmental disorders, repetitive behavior, clinical measures.

Accepted January 24, 2006.

Supported by contracts from the National Institute of Mental Health (N01MH70001 to Dr. McDougle, N01MH70009 to Dr. Scahill, N01MH80011 to Dr. Aman, and N01MH70010 to Dr. McCracken), General Clinical Research Center grants from the National Institutes of Health (M01 RR00750 to Indiana University, M01 RR06022 to Yale University, M01 RR00034 to Ohio State University, and M01 RR00052 to Johns Hopkins University). Korczak Foundation, Dr. Scahill. Study medications were donated by Janssen Pharmaceuticals. The authors acknowledge the efforts of M. Davies, M.P.H., S. Chuang, Ph.D., J. Robinson, M.E.D., and D. McMahon, M.S., for statistical consultation and data management, and permission to adapt the CYBOCS for PDD by W.K. Goodman, M.D., S.A. Rasmussen, M.D., and L.W. Price, M.D.

The CYBOCS-PDD is available from Dr. Scahill.

From the Research Units on Pediatric Psychopharmacology Autism Network (see end of text for authors' affiliations).

Reprint request to Dr. Lawrence Scahill, Yale Child Study Center, P.O. Box 207900, New Haven, CT 06520; e-mail: lawrence.scahill@yale.edu.

DOI: 10.1097/01.chi.0000220854.79144.e7

Pervasive developmental disorders (PDDs) such as autistic disorder, Asperger's disorder, and PDD not otherwise specified (NOS) are chronic conditions of early childhood onset. These disorders share common features, but also have important differences. For example, the diagnosis of autistic disorder requires the presence of delayed and deviant language, impaired social skills, and repetitive behavior and restrictive interests (*DSM-IV-TR*; American Psychiatric Association, 2000). Asperger's disorder is characterized by restricted interests and impaired socialization, but does not involve significant communication delay. In children with PDD-NOS, repetitive behavior or restricted interests may not be prominent (Walker et al., 2004). Because repetitive behavior is a central feature of the PDDs, it warrants careful assessment in both clinical and research settings.

Repetitive behaviors also occur in other disorders including obsessive-compulsive disorder (OCD) and Tourette's syndrome. Despite obvious similarities, there are striking differences in repetitive behaviors observed in these disorders and PDD. Children with OCD describe recurring thoughts that are difficult to dislodge; rituals are often directed at removing contaminants or harm prevention (Scahill et al., 2003). Children with Tourette's syndrome are more likely to engage in repetitive behavior to achieve a sense of completion rather than harm prevention (Scahill et al., 2003). In OCD or Tourette's syndrome, children describe their recurring thoughts as bothersome and the need to perform repetitive behavior as distressing. By contrast, restricted interests and repetitive behavior may not be bothersome to children with PDD, who may indeed derive pleasure or relief from their repetitive behavior (McDougle et al., 2000a). Nonetheless, some children with PDD, particularly higher functioning children, can be diagnosed with OCD and may engage in repetitive behavior to reduce anxiety (Bodfish et al., 2000; Martin et al., 2003; McDougle et al., 2000b).

Repetitive behavior in PDD may be expressed by hand flapping, spinning objects, ordering or arranging objects, repeating phrases, replaying the same video segment over and over, and preoccupations with specific topics. Left to their own, children with PDD may engage in these behaviors for extended periods of time. Indeed, distress may only be manifest when the child is prevented from continuing the repetitive behavior. The neurobiology of repetitive behaviors in PDD remains unknown, although several neurochemical systems have been implicated (Hollander et al., 2003; McDougle et al., 2000b).

Despite being one of the three core dimensions of PDD, relatively few instruments have been designed to assess repetitive behaviors in this population (Bodfish et al., 1999; Militerni et al., 2002). Accurate assessment of repetitive behaviors may be useful during the diagnostic evaluation of PDD. In addition, prominent repetitive behaviors or restricted interests can interfere with functioning and, therefore, may become the target of treatment. Reliable and valid assessment of repetitive behavior is a prerequisite for accurate baseline measurement and tracking progress in clinical and research settings. The purpose of this study was to examine the reliability, validity, and clinical utility of the Children's Yale-Brown Obsessive Compulsive Scales modified for PDD (CYBOCS-PDD).

METHOD

Setting and Subjects

The sample of 172 medication-free children (145 boys and 27 girls) had a mean age of 8.2 ± 2.6 years (range 5–17). Subjects were enrolled in one of two randomized clinical trials conducted by the Research Units on Pediatric Psychopharmacology (RUPP) Autism Network (one participant did not have a CYBOCS-PDD rating at baseline). The first study was a double-blind, placebo-controlled trial of risperidone in children with autism accompanied by aggression, tantrums, and self-injury (Research Units on Pediatric Psychopharmacology [RUPP], 2002). The second study was a double-blind, placebo-controlled trial of methylphenidate in children with PDD and hyperactivity (Research Units on Pediatric Psychopharmacology [RUPP] Autism Network, 2005). Written informed consent was provided by a parent or guardian and assent was obtained from the child when possible.

Procedures

The Children's Yale-Brown Obsessive Compulsive Scale (CYBOCS) is a semistructured clinician rating that measures the current severity of obsessions and compulsions in youth with OCD (Scahill et al., 1997). It is a slightly modified version of the original adult instrument (Goodman et al., 1989). The CYBOCS begins with a detailed symptom checklist of possible obsessions and compulsions. Obsessions and compulsions are then each rated on a 0 to 4 scale across five severity items: Time Spent, Interference, Distress, Resistance, and Degree of Control. Thus, there are three summary scores: Obsessions (0–20), Compulsions (0–20), and a Total score (0–40).

In preparation for use in clinical trials, RUPP Autism Network investigators modified the CYBOCS for use in children with PDD. First, the Obsessions checklist and severity scales were dropped because of the well-known cognitive and communication limitations in this population. Furthermore, because the ascertainment of obsessions in this population would be impeded by language impairment, the severity scales for obsessions would likely be zero and not informative.

Second, the compulsions checklist was retained from the original CYBOCS and expanded to include repetitive behaviors commonly seen in children with PDD. For example, we added repetitive water play, hand flapping, rocking, and echolalia. Third, given that parents are the primary informant for children with PDD, we made selective alterations on the probes for Compulsion severity items. For example, rather than relying on the child to report the distress associated with the repetitive behavior, the parent was asked to estimate the child's level of distress if the child were to be prevented from performing the repetitive behavior. Similarly, the parent was asked to judge whether the child makes an effort to resist the performance of repetitive behavior. To determine the child's level of control, we asked parents to describe the child's response to their directives to stop the behavior. The wording in the anchor points followed from the assumption that children who immediately return to the repetitive behavior are exerting less control over the behavior than a child who accepts the directive and does

not immediately return to the behavior. As in the original CYBOCS, severity items (Time Spent, Interference, Distress, Resistance, and Degree of Control) are rated from 0 to 4. Thus, scores range from 0 to 20 (instrument and instructions are available on request).

Before the enrollment of subjects in the RUPP Autism Network trials, investigators were systematically trained to administer the modified CYBOCS. An in-person training session described the organization, administration, and scoring of the instrument. This training session was followed by a live or taped demonstration of the interview by an experienced rater (L.S.), which was corated by interviewers in training. This was followed by an in-person or telephone conference during which the repetitive behaviors and severity ratings were reviewed. The interviewers then independently rated four additional taped interviews for reliability. To be considered reliable, raters had to score within 15% of the gold standard rating on the total score for each of these four interviews established by an experienced rater (L.S.). The expert rater did not conduct the videotaped interviews; thus, the gold standard ratings were established under the same conditions used by the other raters. Raters who did not meet this reliability criterion received additional training and were required to rate additional tapes reliably before they were allowed to conduct study interviews. The same training methods were used for new raters. Clinicians who performed ratings in the first study (risperidone versus placebo) were recertified for the second study (methylphenidate versus placebo).

The CYBOCS-PDD interview was administered to primary caretakers at baseline when subjects were medication free and at regular intervals throughout the medication trials. In most instances, the child was present during the assessments, which allowed direct participation in the interview and first-hand observation of behavior.

Aberrant Behavior Checklist (ABC). The ABC is a 58-item informant-based scale with five subscales: I, Irritability (tantrums, aggression and self-injury, 15 items); II, Social Withdrawal (response to others, initiation of interaction, 16 items); III, Stereotypic Behaviors (mannerisms and repetitive movements, seven items); IV, Hyperactivity (16 items); and V, Inappropriate Speech (excessive talking, repeating phrases, four items). These factors have been confirmed in several studies (Aman et al., 1985). The ABC has normative data in developmentally disabled populations (Brown et al., 2002) and is sensitive to change (Research Units on Pediatric Psychopharmacology [RUPP] Autism Network, 2002).

Autism Diagnostic Interview-Revised (ADI-R). The ADI-R is a structured parent interview used to support the diagnosis of autism in children and adults (Lord et al., 1997). Training on the ADI-R is rigorous, involving didactic sessions, supervised administration of a live interview, reliable rating of a taped interview, and demonstrated competence in at least three taped interviews by the new rater.

Child Symptom Inventory (CSI). The 126-item, *DSM-IV*-based checklist rated from 0 to 3 by the primary caretaker. Scores of 2 or 3 are regarded as a positive symptom, and the CSI has good reliability and validity (Gadow and Sprafkin, 1994). To evaluate convergent validity with the CYBOCS-PDD, the sum of CSI Compulsions and Tics scales was used to capture observable rituals and stereotypic behavior.

Vineland Adaptive Behavior Scales. The Vineland is a semi-structured, parent interview that measures the child's competence in communication, daily living skills, and socialization. The scale is a standard assessment in children with developmental disabilities with excellent reliability and validity for each domain (Sparrow et al., 1984).

Intellectual Functioning. Children were assessed on one of several intelligence tests: WISC-III (Wechsler, 1991; 29% of the sample); Leiter International Performance Scale-Revised (Roid and Miller, 1997; 28% of the sample); Mullen Scales of Early Learning (Mullen, 1995; 23% of the sample); Slosson Intelligence Test (Jensen and Armstrong, 1985; 14% of the sample); or Wechsler Preschool and Primary Scale of Intelligence-Revised (Wechsler, 1989; 2% of the sample). Eighteen subjects (11%) could not be tested. Because several different tests were employed, children were classified categorically (e.g., average ability; borderline, mild, moderate, or severe mental retardation).

Statistical Analyses

The first set of analyses examined sample differences across the risperidone and methylphenidate trials. This was followed by a comparison of the CYBOCS-PDD scores across the five research sites.

Reliability of the CYBOCS-PDD was evaluated by computing the intraclass correlation (ICC) across clinical raters who participated in the training sequence before launching these trials and by internal consistency (Cronbach's α). Cronbach's α provides a correlational index that reflects the homogeneity of the individual item scores with the total score (Cronbach, 1951). Item analyses (recalculation of the α value with each item sequentially deleted) were performed to determine whether any of the five severity items detracted from the overall reliability estimate. By convention, ICC values of ≥ 0.75 and α values of ≥ 0.8 are considered excellent. To estimate convergent and divergent validity for the CYBOCS-PDD score, we calculated Pearson correlations for theoretically similar indices (e.g., ADI-R Stereotypy, ABC Stereotypy, CSI combined Compulsions and Tics subscales) and dissimilar measures (ADI-R Social Deficit score, Vineland Adaptive Behavior Scales, ABC Irritability and Hyperactivity subscales).

We also conducted several exploratory analyses. To determine whether more severe levels of repetitive behavior were associated with lower cognitive functioning, we examined the reliability and validity of the CYBOCS-PDD in children with an IQ ≥ 70 (higher IQ group) and an IQ < 70 (lower IQ group). CYBOCS-PDD scores across these two groups were compared by ANOVA. Cronbach's α and discriminant validity were also recalculated within each subgroup.

Logistic regression was used to identify clinical characteristics associated with the probability of being in the highest quartile on the CYBOCS-PDD in this sample (i.e., the 55 subjects with CYBOCS-PDD scores of 17 or greater). First, univariate models were examined for the effects of diagnosis (autistic disorder versus PDD-NOS and Asperger's disorder), IQ (above or below IQ of 70), gender, and age. This set of models was used to determine which covariates belonged in the final model. Second, we employed a hierarchical, stepwise modeling procedure with measures of autism severity entered first, followed by measures of repetitive behavior, third by measures of serious behavioral problems such as aggression and hyperactivity, and finally adaptive behavior. To be included in the final parsimonious model, the variable had to be significant (via Wald test) and had to improve the overall model ($p = .05$).

Because CYBOCS-PDD items 4 and 5 (Resistance and Control) rely on parental inference in children with limited verbal ability and insight, we explored an alternative scoring method using only items 1 through 3. This evaluation included calculation of the reliability and validity of items 1 through 3 (Time Spent, Interference, and Distress, respectively) and the correlation of the three-item and five-item scale across the entire sample (risperidone and methylphenidate studies). To compare the three- and the five-item scales on sensitivity to change with treatment, we examined data from the risperidone versus placebo trial (McDougle et al., 2005 for detailed

description of results). Although baseline methylphenidate data were used for reliability and validity data, only data from the risperidone study were used to examine change.

RESULTS

Most subjects were diagnosed with autistic disorder (N = 152), 6 were diagnosed with Asperger's disorder and 14 with PDD-NOS. Table 1 presents the characteristics of the sample.

TABLE 1
Clinical Characteristics of Subjects in RUPP Autism Network Trials

Variable	No.	%
Study source		
Risperidone study	100	58
Methylphenidate study	72	42
Gender		
Male	145	84
PDD diagnosis		
Autistic disorder	152	88
Asperger's disorder	6	4
PDD-NOS	14	8
Cognitive ability (IQ)		
No score	18	11
Average (≥86)	23	13
Borderline (71–85)	26	15
Mild MR (50–70)	40	23
Moderate MR (36–49)	24	14
Severe MR (21–35)	24	14
Profound MR (≤20)	17	10
Ethnicity/race		
White	120	70
Black	20	12
Asian/Pacific Islander	14	8
Hispanic/Latino	11	6
Other	7	4
	Mean (SD)	
Age, yr	8.2 (2.6)	
Vineland Adaptive Behavior Scales		
Communication	50.76 (20.37)	
Daily Living Skills	43.63 (20.72)	
Socialization	53.35 (16.12)	
Composite	45.15 (17.83)	
Aberrant Behavior Checklist		
Irritability	22.22 (9.85)	
Social Withdrawal	14.71 (8.99)	
Stereotypy	8.92 (5.38)	
Hyperactivity	32.64 (8.94)	
Inappropriate Speech	5.82 (3.92)	

Note: RUPP = Research Units on Pediatric Psychopharmacology; PDD = Pervasive Developmental Disorder; PDD-NOS = Pervasive Developmental Disorder not otherwise specified; MR = mental retardation.

Subjects in the risperidone trial were more impaired on measures of behavioral problems and lower in IQ and on measures of adaptive behavior than those in the methylphenidate trial. The mean CYBOCS-PDD scores were 15.3 ± 3.36 for risperidone subjects versus 13.50 ± 3.79 for methylphenidate subjects (t = 2.59; $p < .05$). Similarly, the risperidone sample had a mean ABC Irritability score of 25.72 ± 7.54 compared with 17.29 ± 10.47 in the methylphenidate sample (t = 6.17; $p < .0001$). Given the entry criteria for these two trials, this heterogeneity was not surprising and was considered desirable for the purpose of evaluating the psychometric properties of the CYBOCS-PDD. Across the five sites, mean values on the CYBOCS-PDD ranged from a lower bound of 11.6 ± 4.75 to a higher bound of 17.0 ± 3.74 ($F_{4,167}$ = 10.24, $p < .01$). However, when the sites with the lowest and highest mean scores were removed from the analyses, the correlations with other measures were consistent with the overall findings. Therefore, data from all of the sites were included in the analyses.

CYBOCS-PDD scores for the 172 children in this sample ranged from 0 to 20, with an average score of 14.4 (± 3.86). As can be seen in Figure 1, the distribution of scores is skewed toward the high end of the scale, with a median of 15 and a modal score of 17. Examination of the average scores across the five items suggests that items 4 and 5 are generally higher with less variability than the first three items (Table 2).

Reliability

The ICC of 0.97 indicated excellent reliability across raters. Cronbach's α coefficient indicated a high level of internal consistency for the CYBOCS-PDD five severity items (α = .85). Examination of the overall α with each of the five items sequentially removed shows that no single item substantially detracts from this internal consistency estimate (Table 2).

Validity

Table 3 shows modest or lower correlations of the CYBOCS-PDD with other measures of repetitive behavior. For example, the correlation with the ADI-R Stereotypy subscale was 0.28 ($p < .0001$); the correlation with the ABC Stereotypy subscale was even lower (r = 0.21, $p < .001$). These r values were similar in magnitude to the correlations of the CYBOCS-PDD scores with the ABC Irritability and Hyperactivity

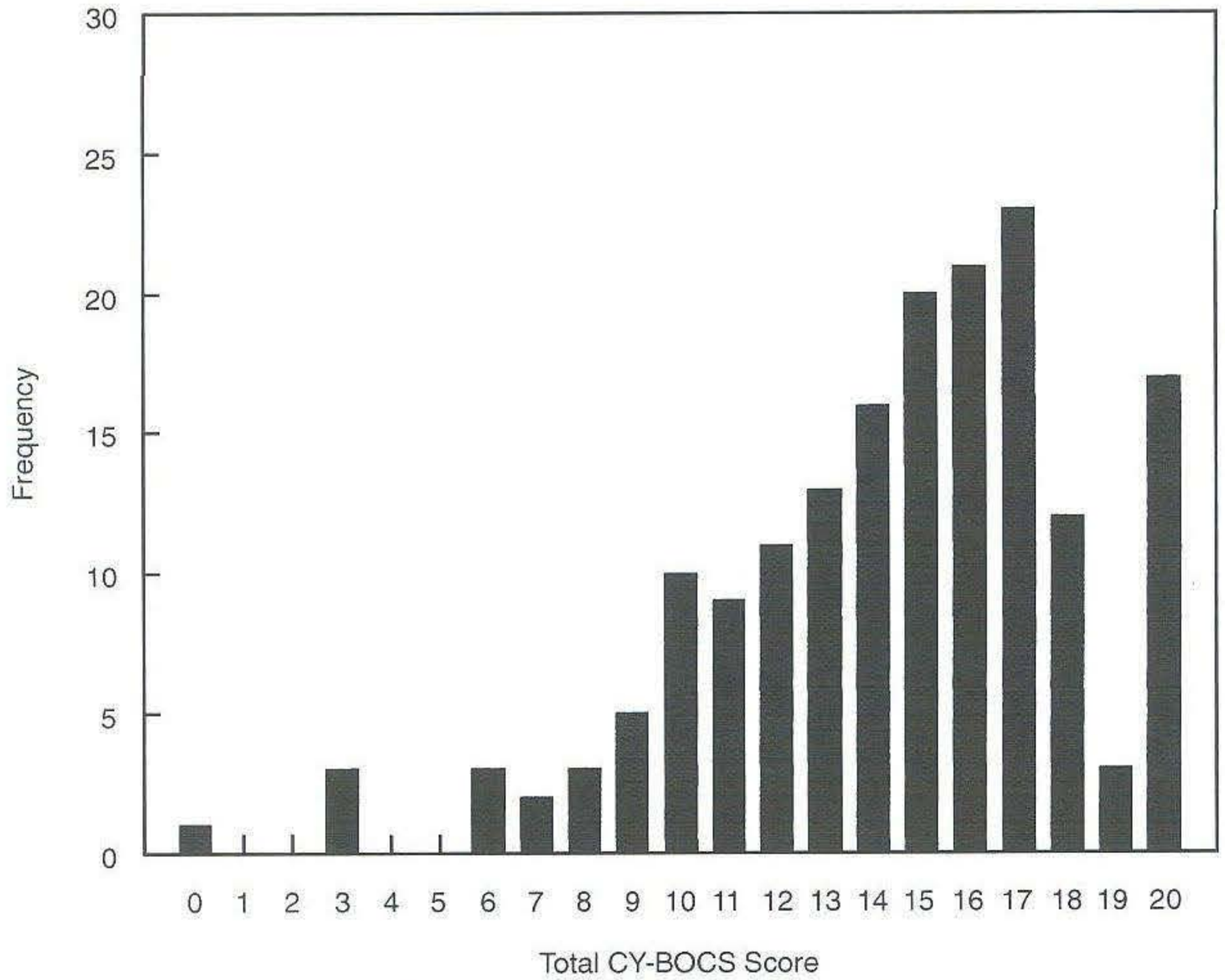

Fig. 1 The distribution of scores on the CYBOCS-PDD (Children's Yale-Brown Obsessive-Compulsive Scales for Pervasive Developmental Disorders) in 172 child participants in RUPP Autism Network trials. The five items are scored from 0 to 4. Higher scores indicate greater system severity; mean score = 14.4 (± 3.86); median = 15; mode = 17; range = 0–20.

subscales ($r = 0.25$ for both). This failure to discriminate between measures of repetitive behavior, measures of behavioral problems, and measures of adaptive skill suggests that the CYBOCS-PDD may be measuring something distinct from these other scales.

Effect of Cognitive Level

When subjects were dichotomized into the two groups (IQ ≥ 70 and IQ < 70), the average CYBOCS-PDD score for the higher IQ group was 14.06 ± 3.54 compared with 14.5 ± 3.98 for the lower IQ group ($t_{170} = 0.69$, NS). Internal consistency, as measured by Cronbach's α, was also similar across groups (α = .81 for the lower IQ group and α = .87 for the higher IQ group). The pattern of convergent and divergent validity, however, differed across the two IQ groups (Table 3). In the higher IQ group, the correlation with the ADI-R Stereotypy was twofold higher than in the lower IQ group ($p = .10$ by Fisher R to Z transformation). The negative correlations between the Vineland and the CYBOCS-PDD shown in Table 3 indicate that children with better adaptive skills (higher Vineland scores) had less severe repetitive behaviors. Although the magnitude of the correlations between the Vineland domains and the CYBOCS-PDD Total score was roughly double in the higher IQ group compared to the lower IQ group, the difference was not significant by Fisher R to Z transformation. Finally, the correlation between the ABC Irritability subscale and the CYBOCS-PDD score was roughly two times greater in the higher IQ group than the lower IQ group ($p = .03$

TABLE 2
Item Analysis and Average Severity Scores on CYBOCS-PDD for Full Sample ($N = 172$)

Item	Mean	SD	α Value With Item Removed
Time spent	2.8	1.00	.821
Interference	2.4	1.00	.814
Distress	2.6	1.05	.813
Resistance	3.3	0.97	.825
Control	3.3	0.84	.825

Note: CYBOCS-PDD = Children's Yale-Brown Obsessive-Compulsive Scales for Pervasive Developmental Disorders.

TABLE 3
Correlations Between CYBOCS-PDD Total Scores for Full Sample and Lower IQ and Higher IQ Groups

	Full Sample (n = 172)		Lower IQ[a] (n = 123)		Higher IQ[a] (n = 49)	
	r	Mean (SD)	r	Mean (SD)	r	Mean (SD)
Repetitive behavior						
ABC Stereotypy Scale	0.21	12.98 (8.54)	0.169	9.83 (5.20)	0.263	6.67 (5.20)
CSI						
Compulsions + tics	0.215**	2.89 (2.44)	0.185*	3.24 (2.54)	0.304*	2.02 (1.90)
ADI-R Stereotypy	0.284**	7.63 (2.66)	0.217	7.62 (2.55)	0.469**	7.65 (2.96)
Maladaptive behavior						
ABC						
Irritability	0.25**	18.14 (11.26)	0.19	20.67 (10.66)	0.39**	12.00 (10.37)
Social Withdrawal	0.327**	14.72 (8.99)	0.288**	14.93 (8.97)	0.428**	14.16 (9.10)
Hyperactivity	0.249**	32.64 (8.94)	0.282**	32.76 (9.04)	0.155	32.33 (8.78)
Inappropriate Speech	0.071	5.82 (3.92)	−0.007	5.62 (4.14)	0.365**	6.29 (3.29)
Autism/adaptive behavior						
ADI-R Social Deficits	0.268**	23.90 (5.05)	0.180*	25.06 (4.18)	0.464**	20.98 (5.87)
Vineland						
Communication	−0.254**	50.76 (20.37)	−0.211*	43.20 (16.03)	−0.456**	69.98 (17.48)
Daily Living Skills	−0.235**	43.63 (20.72)	−0.181*	38.23 (17.63)	−0.367**	57.35 (21.80)
Socialization	−0.272**	53.35 (16.12)	−0.198*	49.49 (15.42)	−0.517**	63.17 (13.60)
Composite	−0.269**	45.15 (17.83)	−0.209*	40.35 (14.87)	−0.426**	57.60 (18.99)

Note: ABC = Aberrant Behavior Checklist; CSI = Child Symptom Inventory; ADI-R = Autism Diagnostic Interview–Revised.
[a] Lower IQ, <70; higher IQ, >70.
* $p \leq .05$; ** $p \leq .01$.

by Fisher R to Z transformation). These results suggest that the CYBOCS-PDD performs somewhat differently for children in the normal IQ range compared with those in the mentally retarded range, but most differences were not significant.

The logistic regression analysis indicated that diagnosis (Wald = 3.30, NS), IQ (Wald = 1.75, NS), gender (Wald = 1.37, NS), nor age (Wald = 0.09, NS) were associated with being in the highest quartile on the CYBOCS-PDD Total (group with score of ≥17). Thus, the stepwise regression procedure did not include any of these variables in the model. The bivariate analyses also examined other measures of repetitive behavior: ADI-R Stereotypy subscale (Wald = 15.38, $p < .001$); ABC Stereotypy subscale (Wald = 6.84, $p < .05$); and CSI composite score (Wald = 8.79, $p < .01$). Because the ADI-R Stereotypy subscale showed the strongest association with the CYBOCS-PDD in the bivariate analysis across these repetitive behavior scales, it was retained in the logistic regression procedure.

The first variable entered in the hierarchical model, the ADI-R Social Deficits score (a measure of autism severity) was significant (Wald = 5.09, $p < .05$). The addition of the ADI-R Stereotypy score improved the overall model fit ($\chi^2 = 16.902$, $p < .001$). Adding the third block of variables, ABC Hyperactivity and Irritability subscale scores (measures of maladaptive behavior), improved the model still further ($\chi^2 = 11.43$, $p < .05$). The final variable entered, the Vineland Adaptive Behavior composite, also improved the model ($\chi^2 = 11.40$, $p < .01$). Although the overall model improved with the addition of these variables, individual variables fell below significance as new variables were introduced. For example, when ADI-R Stereotypy was added to the model, the ADI-R Social Deficit was no longer significantly associated with the highest quartile on the CYBOCS-PDD. Likewise, in the presence of the other variables, the ABC Irritability subscale did not contribute uniquely to the model (Wald = 1.42, not significant). Therefore, the most parsimonious model included ADI-R Stereotypy (Wald = 14.05, $p < .01$), ABC Hyperactivity (Wald = 9.34, $p < .01$), and the Vineland Adaptive Behavior Composite (Wald = 12.38, $p < .01$). This model was significant ($\chi^2 = 17.78$, $p < .05$), correctly classified 78% of the children in the highest quartile on the CYBOCS-PDD, and explained only about 23% of the variance, suggesting intercorrelation among excluded variables.

Five Severity Items versus Three Severity Items

Exploratory analyses of the three-item scale (Time Spent, Distress, and Interference) versus the full five-item scale (including Resistance and Control) showed nearly identical values for internal consistency (α = .83 for items 1–3 compared to .85 for the 5 items). Thus, although there was a suggestion that items 4 and 5 showed less variability across this sample (Table 2), there was no appreciable increase in internal consistency in the alternative scoring method using the three-item scale. In addition, the correlation between the three- and the five-item scales was 0.93, also suggesting no important differences between the two versions of the scale.

In the double-blind phase of the RUPP Autism Network risperidone trial, the five-item CYBOCS-PDD demonstrated sensitivity to change (McDougle et al., 2005). Using the baseline score as a covariate, ANCOVA on the change to endpoint showed no difference in sensitivity to detect change for the first three items (Time Spent, Interference, Distress) versus the full five-item scale (including the Resistance and Control items). The mean score on the five-item scale for the entire sample (N = 100) at the baseline in the risperidone trial was 15.3 ± 3.36. The risperidone group went from 15.5 ± 2.73 at baseline to 11.7 ± 4.02 at endpoint compared with 15.2 ± 3.88 to 14.2 ± 4.81 for the placebo group (ANCOVA = $F_{1,78}$ = 11.66, p = .001). For the three-item scale, the mean for whole group (N = 100) at baseline was 8.5 ± 2.86. The risperidone group went from 8.4 ± 2.10 at baseline to 5.6 ± 2.55 at endpoint compared with 8.5 ± 2.58 to 7.9 ± 3.00 for placebo (ANCOVA = $F_{1,78}$ = 17.05; $p < .001$).

DISCUSSION

The RUPP Autism Network modified the CYBOCS to measure the severity of repetitive behavior in children with PDD before and after treatment. To evaluate this modified instrument, we examined baseline information from 172 medication-free subjects from one of two RUPP Autism Network placebo-controlled trials. Scores on CYBOCS-PDD ranged from 0 to 20, but this range is potentially misleading given the mean score of 14.4, median of 15, and a modal score of 17. Whether this lack of variability is related to this instrument when applied to children with PDD or a reflection of these samples is unclear. Nonetheless, the observation that the CYBOCS-PDD was able to detect change in the placebo-controlled risperidone study suggests that a wider range of scores on this measure is possible in children with PDD.

The training procedures used in the RUPP Autism Network trials achieved excellent interrater reliability (ICC = 0.97). Visual inspection suggests less variability for the Resistance and Control scales (items 4 and 5) compared with the first three items. However, when each of the five items was removed one at a time, no single item detracted substantially from the internal consistency. Although the CYBOCS-PDD significantly correlated with other measures of repetitive behavior (e.g., ADI-R Stereotypy score and CSI Compulsion-Tic scale), these correlations were modest and similar in magnitude to the correlations on measures of maladaptive behavior.

These modest correlations and the failure of the CYBOCS-PDD to discriminate between measures of repetitive behavior and maladaptive behavior suggests that it may be measuring something distinct from the other scales used in these clinical trials. The CYBOCS-PDD asks the parent to specify the child's repetitive behavior from a list of possible behaviors and to provide information on severity. Thus, all of the identified repetitive behaviors, including stereotypic movements, more complex rituals, or circumscribed interests, may contribute to the CYBOCS-PDD total score. Given that the other measures (ABC Stereotypy, ADI-R Stereotypy, and CSI Compulsions and Tics) include a limited set of mostly stereotypic behavior, their moderate correlations with the CYBOCS-PDD score are not surprising. For example, one of the items on the ABC Stereotypy subscale asks the parent to rate the item "stereotyped behavior; or abnormal, repetitive movement." It seems unlikely that a parent would consider a child's compulsive replaying of a cartoon videotape over and over again as a stereotypic movement.

Given the potentially wide range of repetitive behavior in children with PDD, some investigators recommend the use of multidimensional measures. Bodfish et al. (1999) classified repetitive behavior into one of six dimensions: stereotyped behavior, self-injurious behavior, compulsive behavior, ritualistic behavior, insistence on sameness, and restricted interests. These conceptually derived factors have been incorporated

into a 39-item parent rating called the Repetitive Behavior Scale. Subsequent analysis proposed a five-factor approach to this scale in which ritualistic behavior and sameness were combined (Lam, 2004). The dimensional approach in the Repetitive Behavior Scale, especially with the validation from factor analysis, appears useful for documenting the phenomenology of repetitive behavior in children with PDD. It may also be useful in identifying genotype–phenotype relationships. A potential drawback of such multidimensional scales, however, is that it may not be valid to collapse across these factor scores. A subject who is high on one scale may be low on another scale. Therefore, dimensional scales may not provide a single index that can be used to evaluate treatment response. The value of dimensional versus more global ratings is also debated among OCD investigators (Leckman et al., 2001).

Another potential strength of the CYBOCS-PDD is that it is a clinician-rated measure. Unlike a parent-rated measure, the CYBOCS-PDD integrates parent report, observation during the evaluation, and, to the extent possible, child report to make severity ratings. Reliable and valid clinician-rated scales to measure change in children with PDD are few in number and recommended for clinical trials (Scahill and Lord, 2004). The data presented here and the findings of McDougle et al. (2005) suggest that the CYBOCS-PDD is stable in the absence of treatment (in the placebo group) and sensitive to change with treatment (risperidone group). Thus, the CYBOCS-PDD appears to be a useful measure of change for repetitive behavior in this population.

In this sample, children with more severe symptoms on the CYBOCS-PDD showed higher scores on maladaptive behaviors (ABC Hyperactivity and Irritability subscales) and lower scores on adaptive functioning (Vineland domains). Although this pattern appeared clearer in the higher functioning children (IQ $\geq$ 70) compared to the lower IQ group, most differences across IQ groups were not significant.

There were hints during our training procedures that the Resistance and Control items may not apply in this population. In the current analyses, there was indeed less variability on these items compared with the other three items. Furthermore, a score of 4 on the Resistance item ("doesn't resist at all") often corresponded with a score of 4 on the Control item ("has no control over the behavior"). However, internal consistency was no different in the three-item versus the five-item scale and both scales were sensitive to change. An obvious disadvantage of the three-item scale is the reduced range. Therefore, we retained the five-item scale and refined the description of the anchor points for the Resistance and Control items in the current version of the CYBOCS-PDD. These minor modifications should provide better guidance to the clinician for scoring these items.

In contrast to children with OCD, children with PDD may not be distressed by their repetitive behavior. Indeed, distress may only become manifest when the child with PDD is thwarted from performing repetitive behaviors. Children with OCD are often secretive about their ritualistic behavior. By contrast, children with PDD often seem unconcerned about the social implications of their behavior. A child who is unaware or uninterested in the social implications of their repetitive behaviors and performs them in public places is arguably more impaired than a child who shows restraint in public places. An additional severity item based on this concept is under investigation. If this item proves to be reliable, it may be a useful addition to the scale. Nonetheless, the current CYBOCS-PDD appears reliable, valid, and applicable in clinical and research settings.

Limitations

This sample was drawn from two large-scale, multisite treatment trials. Although this procedure provided a large, well-characterized clinical sample, the children with PDD had serious behavioral problems. Moreover, the relatively high CYBOCS-PDD scores notwithstanding, the subjects in these trials were not selected for the presence of interfering repetitive behavior. Thus, the results may not generalize to children with PDD who have less severe problem behaviors or those with prominent repetitive behavior as a primary problem. Another limitation is that several IQ tests were used, which dictated the use of categorical classification of IQ and restricted our ability to evaluate the relationship of IQ and repetitive behavior as measured by the CYBOCS-PDD.

Clinical Implications

The CYBOCS-PDD is a clinician-rated instrument that can be administered relatively quickly. It appears to measure repetitive behaviors that are not captured on

parent-rated measures such as the ABC Stereotypy subscale. It also provides an easily interpreted index of severity for repetitive behaviors that is not subject to random fluctuation, but is sensitive to change with treatment. When administering the CYBOCS-PDD, it is useful to have the child present to integrate clinical observation into the rating. For example, a child may engage in a repetitive behavior that may not have been endorsed by the parent who has become accustomed to the child's repetitive behaviors. In addition, because the CYBOCS-PDD often relies on parental inference, it is often useful to ask the parent to illustrate their inferences with specific examples. For verbal children who describe unwanted obsessive thoughts, the original CYBOCS, which includes the Obsessions scales, could be informative in such cases (Scahill et al., 1997).

MEMBERS OF THE RESEARCH UNITS ON PEDIATRIC PSYCHOPHARMACOLOGY AUTISM NETWORK

Ohio State University: Principal Investigator Michael G. Aman, Ph.D., Co-investigators L. Eugene Arnold, M.Ed., M.D., Yaser Ramadan, M.D., Andrea Witwer, B.S., Ronald Lindsay, M.D., Kristen S. L. Lam, Ph.D.; Indiana University: Principal Investigator Christopher J. McDougle, M.D., Co-investigators David J. Posey, M.D., Naomi Swiezy, Ph.D., Arlene Kohn, B.A.; University of California at Los Angeles: Principal Investigator James T. McCracken, M.D., Co-investigators Bhavik Shah, M.D., Pegeen Cronin, Ph.D., James McGough, M.D., Lisa Sea-Yun Lee, B.A.; Yale University: Principal Investigator, Lawrence Scahill, M.S.N., Ph.D., Co-investigators Andres Martin, M.D., Kathleen Koenig, M.S.N., Deirdre Carroll, M.S.N., Christopher Young, M.D., Allison Lancor, B.A., Anastasia Dimitropoulos, Ph.D., Susan K. Williams, Ph.D.; Kennedy Krieger Institute: Principal Investigator Elaine Tierney, M.D., Co-investigators Jaswinder Ghuman, M.D., Nilda M. Gonzalez, M.D., Marco Grados, M.D.; National Institute of Mental Health: Principal Investigator Benedetto Vitiello, M.D., Co-investigator Louise Ritz, M.B.A.; Columbia University (statistical analysis): Shirley Chuang, Ph.D., Mark Davies, M.P.H.; Nathan Kline Institute (data management): James Robinson, M.E.D., Don McMahon, M.S. Anastasia Dimitropoulos, Ph.D., is with Case Western Reserve University.

Disclosure: Dr. Scahill has affiliations with Janssen, Pfizer, and Bristol-Myers Squibb. Dr. Aman has affiliations with Janssen, Eli Lilly, Forest Laboratories, and Abbott. Dr. Arnold has affiliations with Eli Lilly, McNeil, Novartis, Noven, Shire, Sigma Tau, and Targacept. Dr. McDougle has affiliations with AstraZeneca, Bristol-Myers Squibb, Eli Lilly, Janssen, PediaMed Pharmaceuticals, and Pfizer. Dr. McCracken has affiliations with Janssen, Eli Lilly, Abbott, Bristol-Myers Squibb, Shire, Wyeth, Pfizer, Cephalon, and McNeil. Dr. Posey has affiliations with Eli Lilly, Pfizer, Forest and Janssen. The other authors have no financial relationships to disclose.

REFERENCES

Aman MG, Singh NN, Stewart AW, Field CJ (1985), Psychometric characteristics of the Aberrant Behavior Checklist. *Am J Ment Retard* 89:492–502

American Psychiatric Association (2000), *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR).* Washington, DC: American Psychiatric Association

Bodfish JW, Symons FW, Lewis MH (1999), *The Repetitive Behavior Scale. Western Carolina Center Research Reports.* Morganton, NC: Western Carolina Center

Bodfish JW, Symons FJ, Parker DE, Lewis MH (2000), Varieties of repetitive behavior in autism: comparisons to mental retardation. *J Autism Dev Disord* 30:237–243

Brown EC, Aman MG, Havercamp SM (2002), Factor analysis and norms for parent ratings on the Aberrant Behavior Checklist—community for young people in special education. *Res Dev Disabil* 23:45–60

Cronbach LJ (1951), Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297–334

Gadow KD, Sprafkin J (1994), *Child Symptom Inventories Manual.* Stony Brook, NY: Checkmate Plus

Goodman WK, Price LH, Rasmussen SA (1989), The Yale-Brown Obsessive Compulsive Scale: I. Development, use, and reliability. *Arch Gen Psychiatry* 46:1006–1011

Hollander E, Novotny S, Hanratty M (2003), Oxytocin infusion reduces repetitive behaviors in adults with autistic and Asperger's disorder. *Neuropsychopharmacology* 28:193–198

Jensen JA, Armstrong RJ (1985), *Slosson Intelligence Test for Children and Adults.* East Aurora, NY: Slosson Educational Publications

Lam KSL (2004), *The Repetitive Behavior Scale-Revised: independent validation and the effects of subject variables.* Doctoral dissertation, Ohio State University, Columbus

Leckman JF, Zhang H, Alsobrook JP, Pauls DL (2001), Symptom dimensions in obsessive-compulsive disorder: toward quantitative phenotypes. *Am J Med Genet* 105:28–30

Lord C, Pickles A, McLennan J (1997), Diagnosing autism: analyses of data from the Autism Diagnostic Interview. *J Autism Dev Disord* 27:501–517

Martin A, Koenig K, Anderson GM, Scahill L (2003), Low-dose fluvoxamine treatment of children and adolescents with pervasive developmental disorders: a prospective, open-label study. *J Autism Dev Disord* 33:77–85

McDougle CJ, Scahill L, McCracken J (2000a), Research units on Pediatric Psychopharmacology (RUPP). Autism Network: background and rationale for the initial controlled study of risperidone. *Child Psychiatry Clin N Am* 9:201–224

McDougle CJ, Kresch LE, Posey DJ (2000b), Repetitive thoughts and behavior in pervasive developmental disorders: treatment with serotonin reuptake inhibitors. *J Autism Dev Disord* 30:427–435

McDougle CJ, Scahill L, Aman MG (2005), Risperidone for the core symptom domains of autism: results from the RUPP Autism Network Study. *Am J Psychiatry* 162:1142–1148

Militerni R, Bravaccio C, Falco C, Fico C, Palermo MT (2002), Repetitive behaviors in autistic disorder. *Eur Child Adolesc Psychiatry* 11:210–218

Mullen E (1995), *The Mullen Scales of Early Learning.* Circle Pines, MN: American Guidance Service

Research Units on Pediatric Psychopharmacology (RUPP) Autism Network (2002), Risperidone in children with autism and serious behavioral problems. *N Engl J Med* 347:314–321

Research Units on Pediatric Psychopharmacology (RUPP) Autism Network (2005), Randomized, controlled, crossover trial of methylphenidate in pervasive developmental disorders with hyperactivity. *Arch Gen Psychiatry* 62:1266–1274

Roid GH, Miller LJ (1997), Leiter International Performance Scale-Revised: examiner's manual. In: *Leiter International Performance Scale-Revised,* Roid GH, Miller LJ . Wood Dale, IL: Stoelting

Scahill L, Kano Y, King RA (2003), Influence of age and tic disorders on obsessive-compulsive disorder in a pediatric sample. *J Child Adolesc Psychopharmacol* 13:S7–S17

Scahill L, Lord C (2004), Subject selection and characterization in clinical trial in children with autism. *CNS Spectr* 9:22–32
Scahill L, Riddle M, McSwiggin-Hardin M (1997), Children's Yale-Brown Obsessive Compulsive Scale: reliability and validity. *J Am Acad Child Adolesc Psychiatry* 36:844–852
Sparrow S, Balla D, Cichetti D (1984), *The Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Service
Walker DR, Thompson A, Zwaigenbaum L (2004), Specifying PDD-NOS: a comparison of PDD-NOS, Asperger syndrome, and autism. *J Am Acad Child Adolesc Psychiatry* 43:172–180
Wechsler D (1989), *Manual for the Wechsler Preschool and Primary Scale for Children-Revised*. San Antonio, TX: The Psychological Corporation
Wechsler D (1991), *Manual for the Wechsler Intelligence Scale for Children-Third Edition*. San Antonio, TX: The Psychological Corporation

---

Systematic Review of Prevalence Studies of Autism Spectrum Disorders J.G. Williams, J.P.T. Higgins, C.E.G. Brayne

Aim: To quantitatively examine the influence of study methodology and population characteristics on prevalence estimates of autism spectrum disorders. *Methods:* Electronic databases and bibliographies were searched and identified papers evaluated against inclusion criteria. Two groups of studies estimated the prevalence of typical autism and all autism spectrum disorders (ASD). The extent of variation among studies and overall prevalence were estimated using meta-analysis. The influence of methodological factors and population characteristics on estimated prevalence was investigated using meta-regression and summarised as odds ratios (OR). *Results:* Forty studies met inclusion criteria, of which 37 estimated the prevalence of typical autism, and 23 the prevalence of all ASD. A high degree of heterogeneity among studies was observed. The overall random effects estimate of prevalence across studies of typical autism was 7.1 per 10,000 (95% CI 1.6 to 30.6) and of all ASD was 20.0 per 10,000 (95% CI 4.9 to 82.1). Diagnostic criteria used (ICD-10 or DSM-IV versus other; OR = 3.36, 95% CI 2.07 to 5.46), age of the children screened (OR = 0.91 per year, 95% CI 0.83 to 0.99), and study location (e.g., Japan versus North America; OR = 3.60, 95% CI 1.73 to 7.46) were all significantly associated with prevalence of typical autism. Diagnostic criteria, age of the sample, and urban or rural location were associated with estimated prevalence of all ASD. *Conclusions:* Sixty-one percent of the variation in prevalence estimates of typical autism was explained by these models. Diagnostic criteria used, age of children screened, and study location may be acting as proxies for other study characteristics and require further investigation. **Archives of Disease in Childhood** 2006;91:8–15.